Cassandra Crossing/ l'IA come proprietà emergente?

(536)—Esiste una scuola di pensiero che ritiene possibile, l'evoluzione "spontanea" dei modelli linguistici in "vera" intelligenza…

Cassandra Crossing/ l'IA come proprietà emergente?

Figure 1:

(536)—Esiste una scuola di pensiero che ritiene possibile, l'evoluzione "spontanea" dei modelli linguistici in "vera" intelligenza artificiale. Di cosa esattamente si parla?

14 aprile 2023— "*I Modelli Linguistici pre-Allenati potrebbero manifestare l'Intelligenza Artificiale come proprietà emergente*".

E' un concetto interessante in cui Cassandra non si era mai imbattuta. Ha invece scoperto che l'ipotesi è, pur in un ambiente molto ristretto, oggetto di dibattito filosofico ed anche scientifico. Molto interessante, e vediamo di spiegare il perché.

Perdonate però la vostra profetessa preferita, perché sono necessarie un bel po' di premesse. Speriamo che almeno i 24 indomiti lettori mi seguano senza esitare.

"*Modello Linguistic*o" è qui da intendersi in senso molto preciso e ristretto, come un programma in grado di generare sequenze di simboli, lettere e parole che rispettino la grammatica, la sintassi, lo stile ed altre caratteristiche di un linguaggio naturale.

GPT-3, GPT-4 e ChatGPT sono software che realizzano Modelli Linguistici, cioè che riescono a generare output in linguaggio naturale in risposta a input in linguaggio naturale. Output che posseggono tutte le caratteristiche del linguaggio naturale, tranne la semantica.

In tutto questo la semantica, cioè il significato di una sequenza di parole in rapporto alla realtà extralinguistica non c'entra assolutamente nulla , non è considerata, non è trattata.

I Modelli Linguistici non trattano "per progetto" questo livello di informazione; detto in termini banali, non "comprendono" e non "conoscono".

Se i Modelli Linguistici fossero realizzati in forma deterministica, fossero quindi "algoritmi", si potrebbe conoscere esattamente cosa possono fare e cosa no, ad esempio se possono fare il riassunto di un testo, oppure se possono tradurre un linguaggio in un altro, ed infine se possono comprendere il significato di una domanda, e fornire sempre la risposta esatta.

Ma così non è. La ricerca sui modelli linguistici ha collezionato sostanzialmente solo insuccessi fino a quando non si è orientata verso Modelli Linguistici di tipo "pre-Allenato". In pratica software molto costosi in termini di risorse consumate per realizzarli, che non erano costruiti con algoritmi deterministici, ma che imparavano da zero un linguaggio, analizzando con algoritmi "statistici" in senso esteso, grandi quantità di esempi in linguaggio naturale.

Un processo equivalente all'allenamento di una Rete Neurale tradizionale.

E oggi questo ha prodotto programmi di grande successo.

Dopo essere rimasti fermi al livello "intellettuale" di Eliza dagli anni '60, negli ultimi 10 anni ci sono stati grandi progressi nella costruzione di Modelli Linguistici Generativi pre-Allenati, ed oggi un Modello Linguistico come GPT può essere "venduto" per qualsiasi impiego, visto che è in grado di simulare in maniera convincente le risposte a qualsiasi domanda. Ma è intelligente quanto Eliza, cioè zero. Lo ripetiamo di nuovo; non è neppure minimamente intelligente, non conosce e non comprende nulla. Non è fatto per questo.

E veniamo alla seconda parte di questa lunghissima introduzione.

Cosa è una "proprietà emergente"? E' un concetto molto comune in fisica, dove compare spesso. E' spiegata abbastanza bene su Wikipedia

"*Nella teoria della complessità il **comportamento emergente** è la situazione nella quale un sistema complesso esibisce proprietà macroscopiche ben definibili, difficilmente predicibili sulla base delle leggi che governano le sue componenti prese singolarmente, scaturendo dunque dalle interazioni lineari e non-lineari tra le componenti stesse[1]: quantunque sia più facilmente riscontrabile in sistemi di organismi viventi o di individui sociali oppure ancora in sistemi economici, diversamente da una credenza oggi diffusa l'emergenza si manifesta anche in contesti molto più elementari, come ad esempio la fisica delle particelle[2] e la fisica atomica[3]. Essa può essere definita anche come il processo di formazione di schemi complessi a partire da regole più semplici, e una esemplificazione può ottenersi osservando il gioco della vita di John Conway, nel quale poche semplici regole fissate per pochi individui di base possono condurre a evoluzioni assai complesse.*"

Un esempio semplice è l'acqua, liquido o vapore omogeneo, trasparente ed incolore, formato da molecole semplici e tutte uguali che, se raffreddata al di sotto dello zero, cristallizza in un solido formando cristalli con bellissime geometrie esagonali. La geometria dei cristalli di ghiaccio e neve è un esempio semplice di proprietà emergente come comportamento collettivo di particelle semplici in risposta ad una condizione esterna

Un altro esempio è l'elio raffreddato fino alla superfluidità, che risale le pareti ed esce spontanea-

mente dal recipiente che lo contiene per “cadere” più in basso sul pavimento, manifestando a livello macroscopico un effetto tunnel.

Analogamente, alcuni filosofi o scienziati che si occupano professionalmente di Modelli Linguistici Generativi pre-Allenati come GPT, ipotizzano che dal comportamento complesso ed assolutamente idiota dei Modelli Linguistici possa prima o poi, non si sa assolutamente come o perché, scaturire una vera “intelligenza” come “proprietà emergente” di nuovi Modelli Linguistici sufficientemente più grandi, più allenati, più complessi.

Insomma, ritengono possibile che un GPT-9000 inizi a rispondere in maniera informata ed esatta alle domande poste in linguaggio naturale, invece di fornire vaneggiamenti ed allucinazioni con qualche raro briciolo di casuale esattezza, che è l’unica cosa che i Modelli Linguistici attuali come GPT sanno fare.

Qualcosa che se chiedessi.

“Quante volte il primi 1000 caratteri di pi greco sono diversi da 1?”

mi rispondesse con grande sicurezza

“*Le prime 1000 cifre di pi greco sono composte da numeri diversi dal numero 1 per la maggior parte delle volte. Infatti, la cifra 1 appare solo in circa 166 volte nella prima 1000 cifre di pi greco. Ciò significa che le cifre diverse dal numero 1 compaiono nell’ordine di 834 volte.*”

ma senza sbagliare di grosso (gli “1” sono esattamente 116) fatto che sta in bella mostra su una pagina di Wikipedia, da lui interamente digerita ma, come dicevamo, non memorizzata e non capita.

E che non dica mai che la Luna è fatta di formaggio verde, ma che per caso, fino ad oggi tutti i LEM e le sonde sono atterrate in punti dove era molto invecchiato e sembrava roccia.

Battute a parte, Cassandra, pur ammettendo candidamente di non essere certo un’esperta di IA, anche se ha letto parecchie cose a riguardo, pure qualche testo di premi Nobel che si sono occupati di teoria della complessità (pare che citare premi Nobel renda più credibile ciò che uno scrive), non è per niente convinta e per niente rassicurata da questa ipotesi. E ci sono due motivi ben distinti.

Il primo motivo di preoccupazione per Cassandra è:

Perché, se una proprietà emergente emergesse davvero da non si sa bene dove per una causa che oggi ignoriamo, **dovrebbe essere proprio quella di saper rispondere in maniera completa, corretta ed esaustiva ad una domanda posta in linguaggio naturale**?

Un Modello Linguistico dotato di una proprietà emergente potrebbe rispondere “42” a qualsiasi domanda.

Oppure potrebbe rispondere in maniera corretta, ma anche calcolata per spingere l’umanità, senza che essa se ne accorga, verso una guerra atomica (Skynet) o verso il benessere universale (Le Macchine di Asimov in “Conflitto evitabile”).

Oppure potrebbe manifestare altre proprietà sgradite, pericolose o semplicemente inutili, che potrebbero comunque restare oltre la nostra capacità di percezione o di comprensione.

Il secondo motivo di preoccupazione di Cassandra è di metodo; la ricerca scientifica fatta a caso, o correndo rischi inutili, non sembra la strada giusta per fare una scienza che porti a reali vantaggi e progressi.

Ai primi fisici nucleari che costruirono la bomba A e poi la H, era venuto il dubbio che l'esplosione potesse innescare una combustione dell'atmosfera che, bruciando ossigeno ed azoto, avrebbe reso inabitabile il pianeta.

Si dice che Fermi stesso facesse questi conti con il suo regolo calcolatore poco prima dell'esperimento Trinity ad Alamogordo, ed è un fatto storico che al fisico nucleare Gregory Breit, fu segretamente affidato proprio il compito di rispondere alla domanda se la bomba A o la bomba H avrebbero provocato questa catastrofe. La risposta fu negativa, e per fortuna dell'umanità, anche esatta. Spesso agli scienziati accade di dare risposte esatte.

Ecco, non è che Cassandra sia più preoccupata di tanto che da GPT nasca Skynet come proprietà emergente, ma piuttosto che si sostenga che un Modello Linguistico possa essere o diventare intelligente, e che per questo motivo l'IA in generale venga pubblicizzata in giro come soluzione immediatamente disponibile a tutti problemi.

Questo perché, qui ed oggi, i Modelli Linguistici vengono usati solo ed esclusivamente in maniera propagandistica, come operazioni finanziarie volte a concentrare soldi e potere sempre più nelle mani di poche entità. Vedi caso, le solite.

Qualora su Github qualcuno dovesse pubblicare il codice sorgente per costruire un "oracolo intelligente" che si possa compilare a casa ed addestrare su una branca specifica di conoscenza, Cassandra sarà lieta e sollevata di essersi sbagliata di grosso, e di dover chiedere scusa ad un sacco di persone.

Ma per ora sente solo i risolini che provengono dall'interno delle false "Intelligenze Artificiali" che vengono introdotte nelle nostre case e nelle nostre vite. Sono proprio come quelli che sentì tanto tempo fa quando tirarono dentro alle mura di Troia quel famoso cavallo di legno.

Per questo, dopo avervi tediato con questa lunghissima esternazione, può solo terminarla con l'usuale esortazione partenopea

"Stateve accuorti"!

Scrivere a Cassandra—Twitter—Mastodon
Videorubrica "Quattro chiacchiere con Cassandra"
Lo Slog (Static Blog) di Cassandra
L'archivio di Cassandra: scuola, formazione e pensiero

By Marco A. L. Calamari on April 17, 2023.

Canonical link

Exported from Medium on August 27, 2025.